From Gaia to PLATO: two missions to understand exoplanets

Author: Clara Giménez Arteaga
Facultat de Física, Universitat de Barcelona, Martí i Franqués 1, 08028 Barcelona, Spain.

Advisor: Josep Manel Carrasco

Abstract: *Context.* PLATO mission will detect terrestrial-type exoplanets orbiting in the habitable zone of solar-like stars and characterise their properties.
Aims. This work analyzes *Gaia* Data Release 2 data with two main goals in mind. First, to study the stellar populations that PLATO will be able to detect in the currently proposed observation fields and evaluate the contribution of sources fainter than the PLATO limiting magnitude, which will contribute to its background level. On the other hand, this work will also allow to update the *Gaia* performances model to be implemented in the *Gaia* Object Generator (GOG).
Methods. We use the *Gaia* Archive interface to obtain the data of the proposed PLATO fields. With the help of Topcat we can analize them. In order to derive PLATO's background and update *Gaia* performances, we reduce the amount of data computing running medians with a Python code, fitting then the results with Gnuplot.
Results. The solar-type stars, F5 to K7 type, interesting for PLATO in the search for exoplanets are the most numerous in both proposed fields of study. The mean background level found in this work is 25.6546 mag/arcsec^2 for SPF and 25.3656 mag/arcsec^2 for NPF. The derived performance model fits better with GDR2's G errors, specially for low G, and will be implemented in GOG.

I. INTRODUCTION

A. PLATO

The PLATO space mission (PLAnetary Transits and Oscillation of stars [1]) will detect and characterise terrestrial planets in the habitable zone of solar-type stars. PLATO will provide a huge database with well determined parameters for all kinds of planetary systems (planetary radii, mean densities, ages, stellar irradiation, and architecture of planetary systems) needed to determine their habitability. PLATO will be launched from Kourou in 2025, into a large amplitude libration orbit around L2.

PLATO's current observation scenario assumes two long-duration observation phases of two different sky fields during 2 years each. The optimal field selection is driven by the number of observable targets and the minimisation of false positives due to crowding. Low interstellar extinction is also preferable in this selection of the field candidates. The currently proposed long-duration fields are:

- Northern PLATO Field (NPF): (l = 65°, b = 30°)
- Southern PLATO Field (SPF): (l = 253°, b =-30°)

It must be noticed that NPF includes part of the Kepler mission field. Both NPF and SPF (Fig.1) are studied in this work, even though they are not definitive and might eventually change.

The PLATO mission is distributed in four stellar samples. We will analyze here only Sample 1 (P1), which is the highest priority objective. It will study stars with spectral types from F5 to K7 and magnitudes down to $G < 11$ mag. This limit is partly set due to planet characterization: determining small planet masses is limited to bright host stars. PLATO will provide accurate and precise bulk properties in this range.

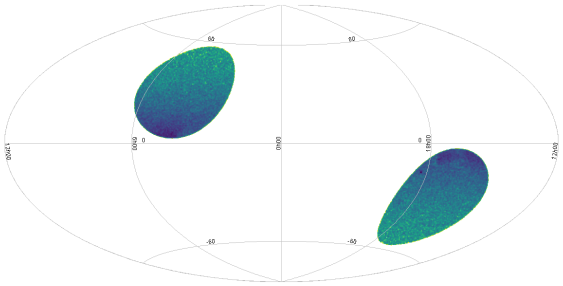

FIG. 1: Aitoff projections of the PLATO fields (NPF left,SPF right) in the Galactic reference frame.

The first aim of this work (Sect. II), is to characterise the observable population of the PLATO fields, to bring light on the target selection process, focusing on P1.

B. Gaia Data Release 2

The ESA *Gaia* space mission was launched in December 2013. The first data release was in 2016, providing measurement of the average flux in the G band for over 1.1 billion sources and parallax measurements for 2.5 million sources [2]. The second *Gaia* data release (GDR2 hereafter) was published in April 2018, providing photometry in the G band for 1.7 billion sources and, for the first time, 1.4 billion sources have $G_{\rm BP}$ and $G_{\rm RP}$ measurements. This new release also contains high-precision parallax and proper motion data for over 1 billion sources [3]. We use data from the *Gaia* catalogue due to its completeness and accuracy in measurements, which will allow us to know better the PLATO fields before its launch.

By analizing the data from GDR2, we will update error models in Sect. IV and calculate the contribution to the sky background of faint sources in Sect. III.

II. POPULATION STUDY

We study here the sources included in GDR2 in the two currently proposed PLATO fields. Except for the first subsection (II A), where we obtain density maps, in this section we only take into account stars with $G < 11$ mag as explained before. NPF has 72378 sources of this kind while SPF has 68316.

A. Density map

As a first step in our study of the PLATO fields, we plot a density map of all sources that are contained in these areas of radius 27° (see Figs. 2 and 3). In order to obtain all the available sources from *Gaia* Archive, we had to download it by small boxes due to the large amount of data. Joining all regions we get the density map of both fields plotted in Figs. 2 and 3, which will be clarifying later when we calculate the sky background.

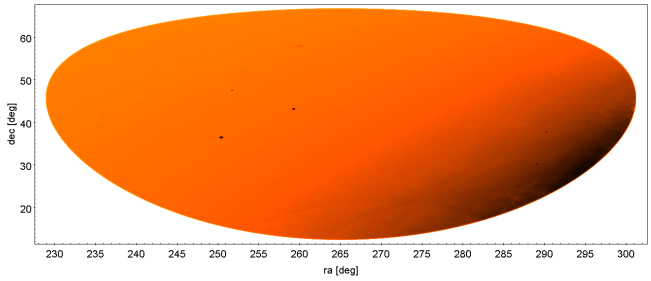

FIG. 2: Density map of the Northern PLATO Field (NPF).

Even though NPF has a galactic latitude of 30°, we can clearly see the increase in density at lower galactic latitudes, where it is closer to the disk of the galaxy. We can identify the globular clusters NGC 6205 at ra=250.42° and NGC 6341 at ra=259.28° [4].

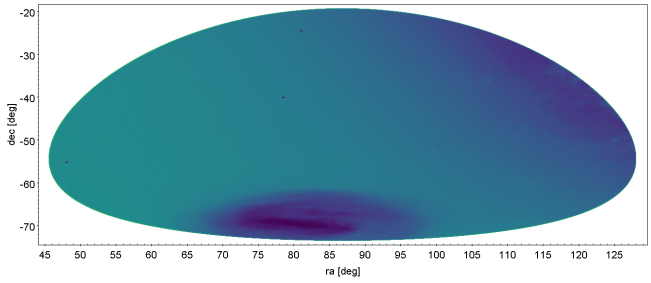

FIG. 3: Density map of the Southern PLATO Field (SPF).

For SPF the galactic latitude is -30° and we can see the density gradient when approaching to the galactic disk (upper right section in Fig. 3). Besides this, we can clearly see that there is a huge density region on the lower part of the map. This spot corresponds to the Large Magellanic Cloud.

B. Magnitudes

GDR2 provides three different photometric mean magnitudes: the *Gaia* unfiltered band (G), which is the widest passband in *Gaia* and therefore the most precise, and the magnitudes obtained using a blue and a red filter ($G_{\rm BP}$ and $G_{\rm RP}$ respectively). Figure 4 shows the number of sources as a function of the magnitude for NPF (the plot for the SPF is very similar), showing, as expected, an increase of the number of sources with the magnitude, being a complete sample for magnitudes with G<11 mag, $G_{\rm BP}$<11.12 mag and $G_{\rm RP}$<10.11 mag.

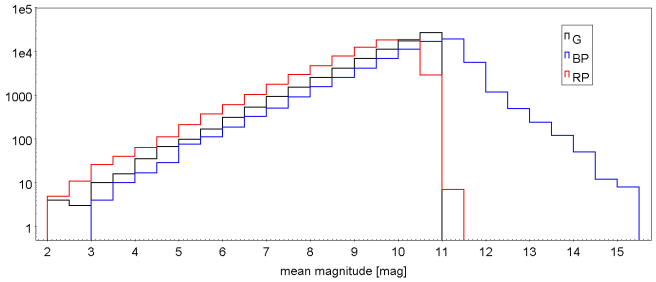

FIG. 4: Histogram of the three mean magnitudes G, $G_{\rm BP}$ and $G_{\rm RP}$ for NPF.

We will use the G magnitude to obtain the absolute magnitude in the H-R diagram. With the other two magnitudes we can get the colour index $G_{\rm BP} - G_{\rm RP}$, which is linked to the temperature of the source, helping in the selection of the targets for PLATO.

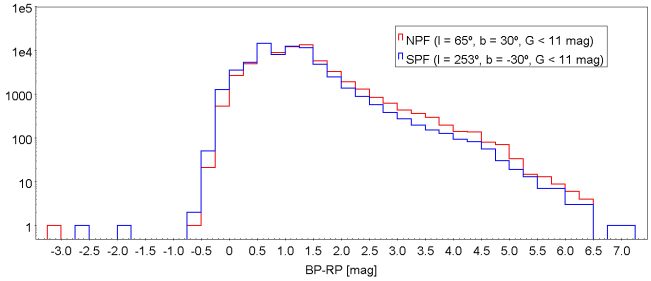

FIG. 5: Colour histogram $G_{\rm BP}$-$G_{\rm RP}$ for NPF and SPF.

We see in Fig.5 that the vast majority of stars have $G_{\rm BP} - G_{\rm RP}$>0. This is due to two phenomena: there are less blue objects, they evolve faster and last less in the sky, and also that we have the interstellar medium that reddens the light, even though at the latitudes of NPF (b=30°) and SPF (b=−30°) this should not be as important as in other regions.

C. Parallax and distance

Parallax gives us information on the distance between us and the measured source. The smaller the parallax angle, the further away the source is [5]. Figure 6 shows the histogram of the GDR2 parallaxes. In both fields, the distribution of parallaxes is rather similar, and the majority of stars have a parallax between 0 and 20 mas.

The *distanceEstimateEdsd* function in Topcat [6] allows us to derive distances from parallax angles. We make use of this function, using a length scale parameter equal to L=1350 pc, as used in [7], and obtain the distance distributions shown in Table I. The closest sources could be useful in the future study of planetary atmospheres.

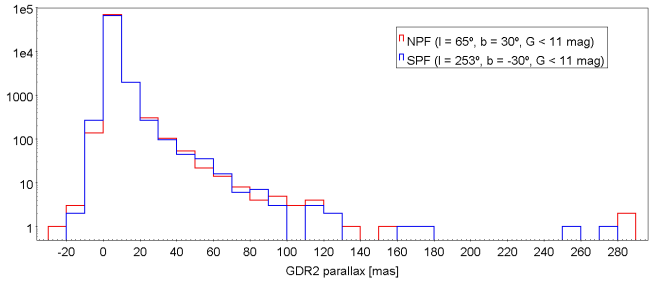

FIG. 6: Histogram of the parallax versus sources for NPF (red) and SPF (blue).

Distance (pc)	NPF		SPF	
	Sources	%	Sources	%
0-100	2474	3	2432	4
100-200	7610	11	7653	11
200-500	23017	32	22096	32
500-1000	19080	26	18227	27
1000-2000	14276	20	12785	19
> 2000	5921	8	5123	7

TABLE I: Number of sources from NPF and SPF at each distance range (in parsecs), and percentage it represents from the total of sources.

D. Variable stars

We can also obtain the population of variable sources of PLATO's fields. This study could be useful to identify potential astrophysical false alarms: due to eclipsing binaries, true transiting planets are sometimes outnumbered [1].

GDR2 variability flag provides the classification in Table II. In the following *Gaia* releases, when more data is collected with also an enhanced processing, this flag will probably add more variable sources than currently present, but now, since most of sources have no available classification on the matter, we expand our knowledge by cross-matching GDR2 with the Simbad catalogue [8]. Simbad classifies variable stars as indicated in Table III, showing only the most populated kinds of variability.

Type	NPF	SPF
Variable	755	533
Not Available	71623	67783

TABLE II: Number of variable sources identified in GDR2.

Type	NPF	SPF
Rotationally Variable Star	584	98
Long-period Variable Star	581	127
Variable Star	433	148
Variable of delta Sct type	208	35
Variable of Mira Cet type	107	56
Eclipsing binary	192	93
Pulsating Variable Star	58	19

TABLE III: Variable sources identified in the SIMBAD catalogue [8].

The *Variable Star* type represents those stars which are known variables but have not been further classified yet. We can see that in NPF the most common are the rotationally variable stars whereas in SPF those are the long-period variable stars.

E. Spectral Classification

The precise determination of colours by *Gaia* allows us to identify bright, nearby cool F-, G-, K-, M-dwarfs and sub-giants across the region covered by the PLATO fields [1]. These stars are PLATO's target in the search for terrestial-type exoplanets, therefore we want to determine the populations of each spectral type. We can associate a spectral type with the effective temperature of the source [5]. This temperature can be determined with the following equation (Carrasco private communication), that relates $T_{\rm eff}$ with the colour $G_{\rm BP} - G_{\rm RP}$ provided by *Gaia* DR2:

$$\theta = 0.44222 + 0.54422(G_{\rm BP} - G_{\rm RP}) - 0.044345(G_{\rm BP} - G_{\rm RP})^2 \quad (1)$$

where $\theta = 5040/T_{\rm eff}$.

With this temperature we can classify the NPF and SPF sources by their spectral type in Table IV. We can see that potential PLATO targets (from spectral types F, G, K and M) are the most numerous. In NPF these sources are the 91% of stars whilst in SPF they are the 87%. We can also see that there is an 11% of solar-type stars (of spectral type G) in both fields. For Sample 1 (P1) we will only study stars from F5 to K7 as explained at the introductory Sect. I.

Type	T_{eff}	NPF		SPF	
		Sources	Fraction	Sources	Fraction
O	$\geq$ 30000 K	1	0%	3	0%
B	10000-30000 K	1571	2%	2805	4%
A	7500-10000 K	4969	7%	5807	9%
F	6000-7500 K	17871	25%	17805	26%
G	5200-6000 K	8011	11%	7276	11%
K	3700-5200 K	33638	46%	30391	44%
M	2400-3700 K	6231	9%	4145	6%

TABLE IV: Stellar classification of NPF and SPF sources due to the spectral type.

F. Hertzsprung-Russell diagram

We can build an observational H-R with the absolute visual magnitude of each star plotted versus its colour index. The absolute magnitude M is given by [5]:

$$M = m + 5(\log(p) + 1) \tag{2}$$

where m is the apparent magnitude and p the parallax (in arcseconds). The colour index $G_{\text{BP}} - G_{\text{RP}}$ is provided in GDR2. The H-R diagram for NPF is shown in Fig.7 (SPF's H-R is very similar). We only plot sources with less than a 20% error in parallax (70879 sources).

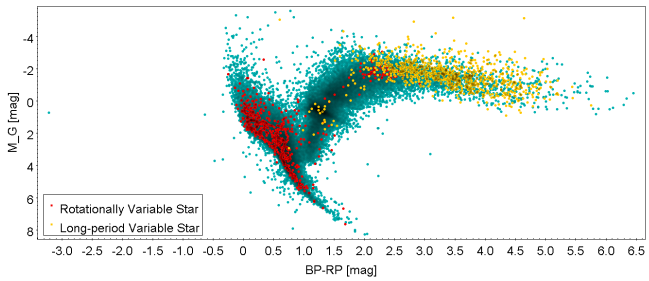

FIG. 7: NPF Hertzsprung-Russell diagram.

Most stars are contained in the main sequence and in the giant's branch (up-right from the MS). There are no visible supergiants (should appear in the extreme upper right-hand corner) neither white dwarfs (lower left of the diagram). NPF's most abundant variable stars (see Table III) are coloured.

III. SKY BACKGROUND

One of the aims of this work is to evaluate the contribution of faint *Gaia* sources to the PLATO background level. The PLATO cameras can detect sources with G<11 mag, fainter sources will contribute to its background. In these calculations we do not take into account the zodiacal contribution of light from the Sun, only faint stellar objects in the PLATO fields, although it would consist in adding an extra component depending on the ecliptic latitude.

PLATO's plate scale is 15.0 arcsec per pixel, for both normal and fast telescopes [1]. The PSF (Point Spread Function) surface is always included within 9 pixels [1], which makes a PSF of 135 arcsec wide. There will be stars inside this PSF size of 135×135 arcsec^2 that will contribute to the light of the source. We want to establish, in an equivalent PLATO's PSF area, the value of the contribution of the total flux emitted by stars which are not detected by PLATO (G>11 mag) but still contribute to the measured signal, as a background.

The background magnitude is given by [5]:

$$G = -2.5 \log\left(\frac{\sum F}{135 \times 135}\right) + 25.7934 \tag{3}$$

where F is the flux, 135 × 135 is PLATO's PSF, and 25.7934 is the zeropoint of the G magnitude from *Gaia* [2]. Figures 8 and 9 show the resulting background levels in NPF and SPF, respectively.

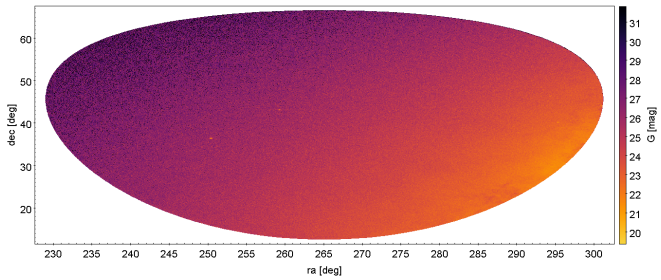

FIG. 8: Background map of the Northern PLATO Field.

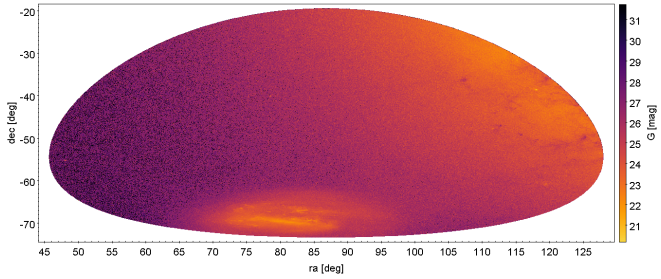

FIG. 9: Background map of the Southern PLATO Field.

Background G	Mean	SD	Minimum	Maximum
NPF	25.3656	2.06809	18.35791	32.19283
SPF	25.6546	2.02195	18.64412	32.25911

TABLE V: Background G values for each PLATO field.

There is a wide interval of magnitudes in both maps. In both cases we can see the relation with the density maps from Sect.II A: wherever there is a higher density of

stars, the apparent magnitude that we receive is smaller (therefore brighter). In SPF the brightest background contribution comes from the Large Magellanic Cloud.

IV. *Gaia* PERFORMANCE

The last aim of this work is to update error models in the *Gaia* Object Generator (GOG) based on GDR2, so we can later adapt them to the PLATO mission. The GOG simulator [9] includes observational errors. We want to update the G error equation that appears on the website [10]. This portal contains a simple performance model for the error of G (σ_G) (see Eq.4), based on the data from GDR1.

$$\sigma_G[\text{mag}] = 10^{-3}(0.04895z^2 + 1.8633z + 0.0001985)^{1/2} \quad (4)$$

where z = MAX$[10^{0.4(12.09-15)}, 10^{0.4(G-15)}]$.

As we will later demonstrate, we can not fit the experimental errors from GDR2 with such a simple equation.

The analysis will be conducted on NPF. We want to find equations that fit the curve $\log(\sigma_G)$ versus G. When facing this issue, we encountered 68.5 million sources to fit. In order to reduce the number of points to be fitted, we calculate the median. Nevertheless, for $G < 7$ mag we still use all data available, as this range is not so populated and the median could be not so representative.

We obtain two main types of fitting equations: polynomial (Eq.5) and Gaussian (Eq.6). The general equations are the following:

$$\log(\sigma_G) = A_0 + A_1\,G + A_2\,G^2 + A_3\,G^3 \quad (5)$$

$$\log(\sigma_G) = y_0 + \frac{B\exp(\frac{-4\ln 2}{w^2}(G-x_0)^2)}{w\sqrt{\pi/4\ln 2}} \quad (6)$$

There are different ranges of G where these equations fit the experimental curve $\log(\sigma_G)$ versus G. Table VI shows the G interval and the corresponding coefficients of the polynomial and gaussian fitting equations.

Figure 10 plots the experimental data $\log(\sigma_G)$ versus G and the ensemble of fitting equations, each in its corresponding G interval.

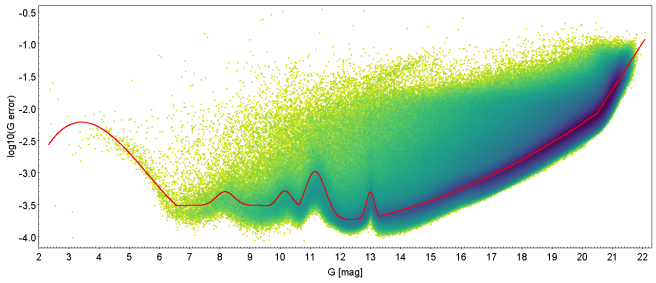

FIG. 10: Error in the G band magnitude.

G		A_0	A_1	A_2	A_3
Min	Max				
0	6.56764	-6.77	3.106036	-0.651712	0.0387
13.284175	20.4988	-2.69652	-0.267082	0.0144829	0
20.4988	22	-17.0177	0.728437	0	0
G		y_0	B	x_0	w
Min	Max				
6.56764	9.640607	-3.51831	0.11032	8.1867	0.6979
9.640607	10.638845	-3.5347	0.09475	10.16084	0.52733
10.638845	12.18945	-3.74274	0.45	11.1657	0.81
12.18945	13.284175	-3.73376	0.1	12.99651	0.32

TABLE VI: Coefficients of the polynomial (Eq.5) and gaussian (Eq.6) equations for each G range.

V. CONCLUSIONS

P1 will target the solar-type stars (F5 to K7), which are the most abundant sources in NPF (82% of the total) and SPF (81%). The mean background level for PLATO found in this work is 25.6546 mag/arcsec^2 for SPF and 25.3656 mag/arcsec^2 for NPF. The background G varies from 18 to 32 mag (per arcsec^2). The derived performance model for the G error fits better with GDR2's data and will be implemented in the *Gaia* simulator.

Acknowledgments

I want to thank my advisor, Josep Manel Carrasco, and my family and friends for their astronomical support.

[1] PLATO Definition Study Report, ESA-SCI(2017)1
[2] D. W. Evans et al., 2018, A&A, Forthcoming article
[3] *Gaia* Collaboration: A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne et al., 2018, A&A, Forthcoming article
[4] F. Bonnarel et al.,"The ALADIN interactive sky atlas", Astron. Astrophys. Suppl. Ser. 143, 33-40 (2000)
[5] Bradley W. Carroll, Dale A. Ostlie, *An Introduction to Modern Astrophysics*,(Pearson Education, 2014,2nd ed.).
[6] M.B. Taylor,"TOPCAT&STIL: Starlink Table/ VOTable Processing Software", Astronomical Data Analysis Software and Systems XIV, Vol. 347 (2005)
[7] Luri et al., 2018, A&A, Forthcoming article
[8] Wenger et al., "The SIMBAD astronomical database", 2000,A&AS,143,9
[9] Luri et al., 2014, A&A 566, A119
[10] https://www.cosmos.esa.int/web/gaia/science-performance